FEBS Letters 584 (2010) 1273–1278

FEBS Letters

journal homepage: www.FEBSLetters.org

Comparative genomic and phylogenetic analysis of vitellogenin and other large lipid transfer proteins in metazoans

Alexander Hayward [a], Tokiharu Takahashi [b,*], William G. Bendena [c,*], Stephen S. Tobe [d,*], Jerome H.L. Hui [a,b,*]

[a] Department of Zoology, University of Oxford, South Parks Road, OX1 3PS, United Kingdom
[b] Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom
[c] Department of Biology, Queen's University, Ontario, Canada K7L 3N6
[d] Department of Cell and System Biology, University of Toronto, 25 Harbord St., Canada M5S 3G5

ARTICLE INFO

Article history:
Received 17 December 2009
Revised 16 February 2010
Accepted 19 February 2010
Available online 24 February 2010

Edited by Takashi Gojobori

Keywords:
Vitellogenin
Apolipoprotein
Microsomal triglyceride transfer protein
Placozoa
Cnidaria
Phylogeny

ABSTRACT

Vitellogenins and other large lipid transfer proteins (LLTP) are well known to play significant roles in the development, metabolism and reproduction of animals. Comparative genomics and phylogenetic analyses of LLTPs using the most comprehensive dataset in metazoans to date are carried out. Our analyses demonstrate that LLTP genes arose significantly earlier, and are more widespread than previously proposed – being present in numerous additional bilaterian and non-bilaterian lineages. A hypothesis is advanced that the most ancestral animal LLTP gene is Vtg, while loss of domains occurred at the bilaterians stem giving rise to apolipoprotein and microsomal triglyceride transfer proteins genes.

1. Introduction

Lipoproteins are crucial components of a wide range of biological processes in animals, from basic cellular structural support to enzymatic and transportation reactions. Some lipoproteins are processed by large lipid transfer proteins (LLTP), which include vitellogenin (Vtg), the cytosolic large subunit of microsomal triglyceride transfer protein (MTP) and apolipoproteins (Apo). Apolipoproteins can be further subdivided into vertebrate apolipoprotein B (ApoB), crustacean apolipocrustacein (ApoCr), and insect apolipophorin II/I precursor (ApoLp-II/I) [1,2].

Over recent decades, extensive effort has gone into identifying the underlying mechanism of LLTP processing in animals (mainly from the phyla Arthropoda and Vertebrata). Today, the primary functions of these LLTPs are relatively well known. For example, ApoB is the primary apolipoprotein responsible for transporting cholesterol to tissues in humans, and high levels of ApoB are related to cardiovascular diseases. MTP is known to be crucial in the processing of lipoprotein, and mutations in MTP cause abetalipoproteinemia in humans. Vtg (Gr: *vitello* = yolk and *gener* = creation) is the most extensively characterized lipoprotein, and plays a central role in vitellogenesis for bilaterians, including arthropods, nematodes and vertebrates [3–5].

Given the relatively high amino acid sequence similarity found among LLTP proteins, surprisingly few studies have investigated their relationships across animal phyla. In addition, most previous studies have assigned cloned gene fragments to particular LLTP gene families on the basis of sequence similarity rather than via phylogenetic analysis. Subsequently, a "chain-reaction" problem in nomenclature arose when a misclassification occurred in a given study. This was recently demonstrated during a study in which phylogenetic trees were constructed using LLTP genes mainly from arthropods and vertebrates, resulting in the long studied "Vtg" gene in crustaceans being re-classified as an apolipoprotein (ApoCr), and the orthologue of insect ApoLp-II/I and vertebrate ApoB [1].

To date, only one LLTP gene has so far been reported in non-bilaterians [10], which hinders the understanding of the LLTP

Abbreviations: Vtg, vitellogenins; LLTP, large lipid transfer proteins; ApoB, apolipoprotein B; ApoLp-II/I, apolipophorin II/I precursor; ApoCr, apolipocrustacein; MTP, microsomal triglyceride transfer protein

* Corresponding authors. Address: Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester, M13 9PT, United Kingdom (J.H.L. Hui).

E-mail addresses: tokiharu.takahashi@manchester.ac.uk (T. Takahashi), william.bendena@queensu.ca (W.G. Bendena), stephen.tobe@utoronto.ca (S.S. Tobe), hui.jerome@gmail.com, jerome.hui@manchester.ac.uk (J.H.L. Hui).

0014-5793/$36.00
doi:10.1016/j.febslet.2010.02.056

family relationships in animals. With the increasing numbers of animal genomes other than vertebrates and insects are becoming available (e.g. [6–9]), it is pertinent to compare the LLTP genes across diverse animal phyla, with the aim of reconstructing an accurate evolutionary history. Here, we have performed the most comprehensive comparative genomic analyses of the LLTP genes in metazoans to date, which lay the framework for further research into the evolution of functions of LLTP genes. Moreover, we propose the first evolutionary model to explain how LLTP genes potentially arose in animals.

2. Materials and methods

2.1. Sequence retrieval

LLTP sequences were obtained in public databases (NCBI and JGI). A dataset consisting of 94 LLTP sequences was assembled for analysis. To avoid confusion arising from different gene nomenclature, we have chosen to follow the terms classified in [1]. Methods, gene names and accession numbers obtained from NCBI and JGI databases are summarized in Supplementary data 7.

Fig. 1. Phylogenetic tree using Maximum-Likelihood method showing the relationships of LLTP genes.

2.2. Phylogenetic reconstruction

All analyses were carried out using amino acid sequences. Sequences were aligned using the program ClustalX [11], with subsequent visual editing. Portions of alignments where information was only available for a small subset of the taxa were excised, resulting in datasets of 349 selected amino acids. Alignments are provided in the Supplementary data 1. Phylogenetic relationships were estimated by Bayesian-Inference and Maximum-Likelihood, with phylogenetic support (clade posterior probabilities/bootstrap support values) displayed on the respective majority rule consensus tree.

Bayesian analyses were carried out in MrBAYES v3.1.2 [12], using the (MC)3 algorithm, with four simultaneous Markov chains

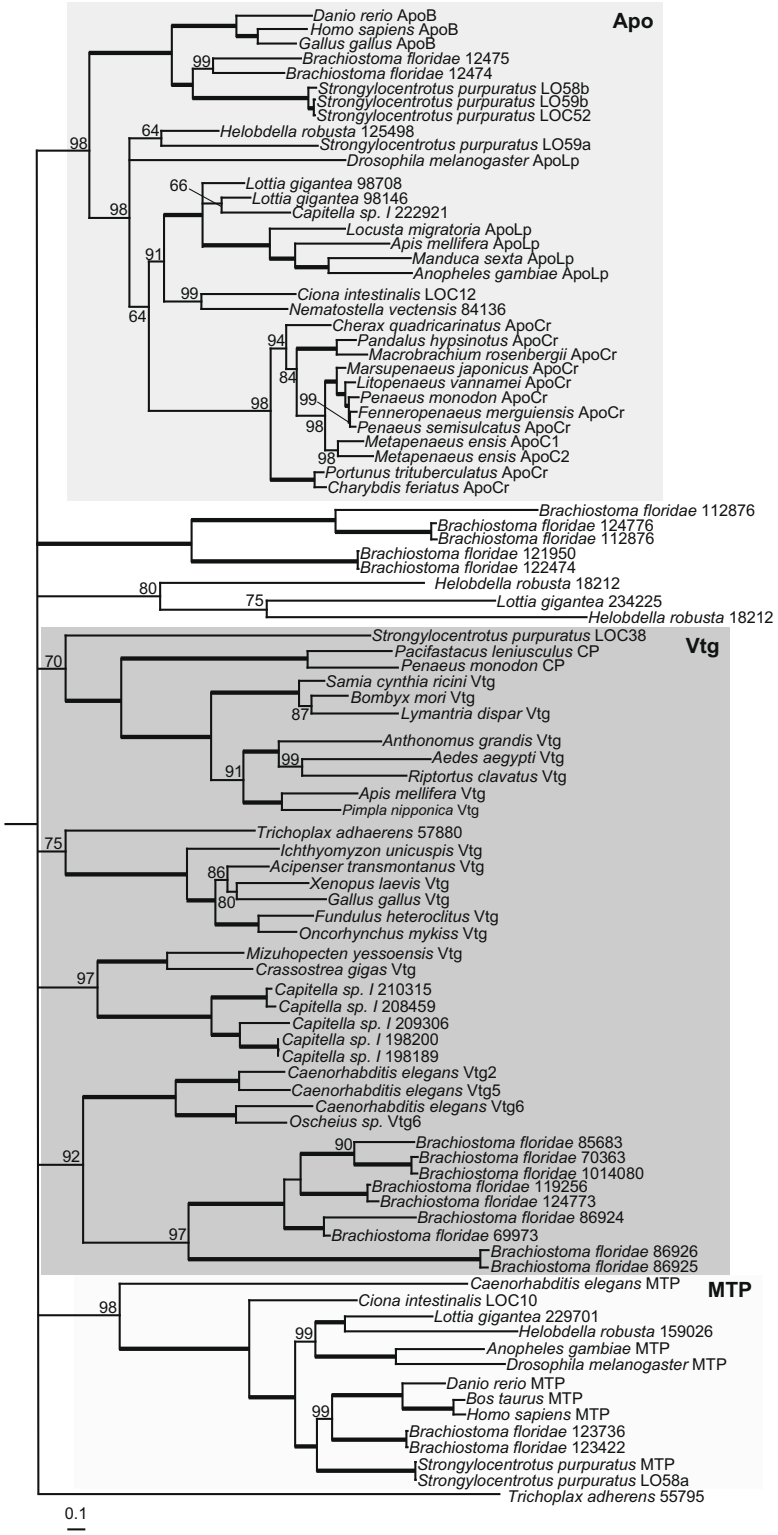

Fig. 2. Phylogenetic tree using Bayesian method showing the relationships of LLTP genes.

	Gene	Tree supported	LPD_N domain	DUF1943 domain	DUF1081 domain	vWD domain
Polychaetes (*Capitella sp.I*)	Apo	B, ML	Csp222921	Csp222921	Csp222921	Csp222921
	Vtg	B, ML	Csp210315	Csp210315		Csp210315
	Vtg	B, ML	Csp198200	Csp198200		Csp198200
	Vtg	B, ML	Csp209306	Csp209306		Csp209306
	Vtg	B, ML	Csp198189	Csp198189		Csp198189
	Vtg	B, ML	Csp208489	Csp208489		Csp208489
Limpets (*Lottia gigantea*)	MTP	B, ML	Lgi229701			
Leeches (*Helobdella robusta*)	MTP	B, ML	Hro159026			
Echinoderms (*Strongylocentrus purpuratus*)	MTP	B, ML	SpuLOC583528			
	Apo	B, ML	SpuLOC583017	SpuLOC583017	SpuLOC583017	
	Apo	B, ML	SpuLOC581779	SpuLOC581779	SpuLOC581779	
	Vtg	B, ML	SpuLOC583638			SpuLOC580672
Cephalochordates (*Brachiostoma floridae*)	Apo	B, ML	Bfl124774	Bfl124774	Bfl124774	Bfl124774
	Apo	ML	Bfl124776	Bfl124776	Bfl124776	
	Vtg	B, ML	Bfl86924			
	MTP	B, ML	Bfl123736			
	MTP	B, ML	Bfl123422			
Tunicates (*Ciona intestinalis*)	MTP	B, ML	CinLOC100183073			

Fig. 3. Table showing the newly identified LLTP proteins in bilaterians. Only LLTP proteins that are supported by both domain classification and at least one phylogenetic method are shown. These animals include polychaetes *Capitella sp. I*, limpets *Lottia gigantea*, leeches *Helobdella robusta*, echinoderms *Strongylocentrotus purpuratus*, cephalochordates *Branchiostoma floridae*, and tunicates *Ciona intestinalis*. "Apo", "MTP" and "Vtg" in the gene column respectively represent the assignment of gene as Apo, cytosolic large subunit of microsomal triglyceride transfer protein, and vitellogenin. "B" and "ML" in the tree column respectively represent the Bayesian and Maximum-Likelihood phylogenetic analyses that support the assignation.

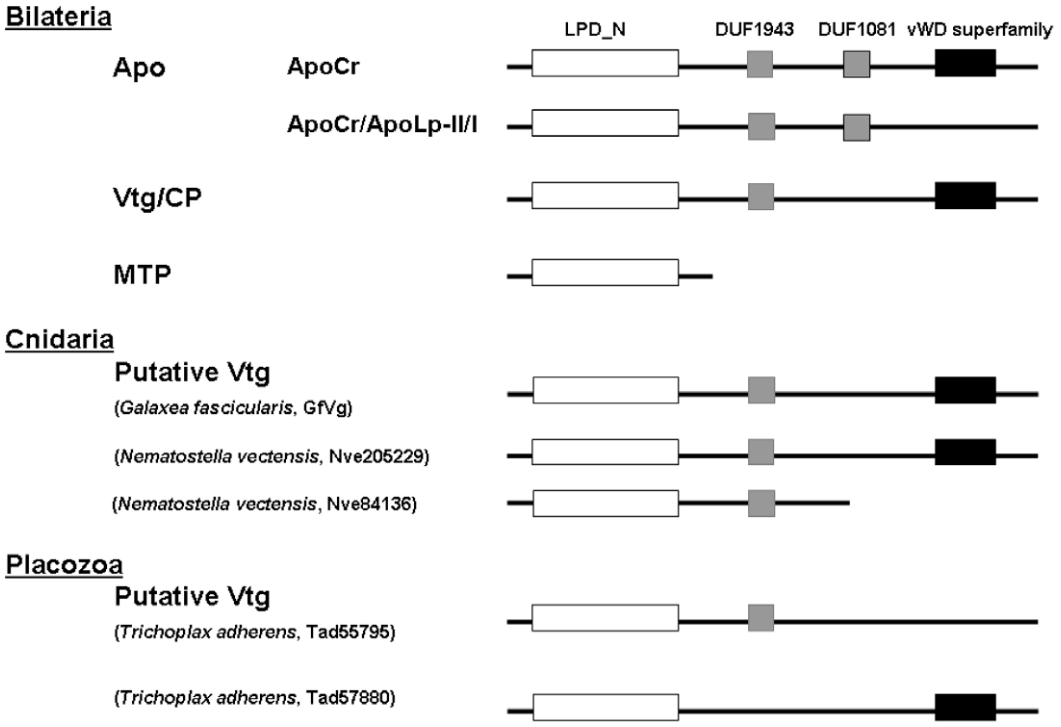

Fig. 4. Schematic summary of domain structures of bilaterian and non-bilaterian LLTP proteins. White boxes (LPD_N) denote the domains of lipoprotein amino terminal region; grey boxes (DUF1943 and DUF1081) denote DUF1943 and DUF1081 domains, and black boxes (vWD superfamily) denote von Willebrand factor type D domains. Note that the diagram is not drawn to scale.

per run (three heated, one cold) and two independent runs per analysis as implemented by default. The model jumping command in MrBAYES was specified, which selects substitution models in proportion to their posterior probability. Chains were run for 1 million generations with a sampling frequency of 100 generations, resulting in 10,000 samples per run. After chain completion, a plot of log-likelihood against sample number was examined manually for each run in order to judge if stationarity had been achieved and to determine the burn-in. Bayesian posterior probabilities were estimated for each clade from the 50% majority rule consensus tree of the sampled trees minus the burn-in. Maximum likelihood analyses were carried out using PhyML v3.0.1 [13] with the substitution models selected by the Akaike information criterion as implemented in Prottest 1.4 [14]. The model selected was LG with gamma distributed rate heterogeneity (G) the dataset. A 50% majority rule consensus tree was constructed from 100 bootstrap pseudoreplicates.

3. Results and discussion

The results of our phylogenetic analyses (Bayesian and Maximum-Likelihood) suggest that bilaterian LLTPs can be generally classified into Vtg, vertebrate ApoB, insect ApoLp-II/I, crustacean ApoCr, and the extracellular cytosolic large subunit of MTP (Figs. 1 and 2). This classification confirms that the genes previously classified as "Vtg" in crustaceans should be reclassified as Apo, and that the crustacean clotting protein (CP) genes are the true Vtg orthologs [1]. By including putative LLTP genes from multiple taxa, our analysis extends this classification by revealing that numerous bilaterian lineages have orthologues of the Apo gene family. The animals represented by these lineages include: polychaetes (*Capitella sp. I*), sea urchins (*Strongylocentrotus purpuratus*), and cephalochordates (*Brachiostoma floridae* (Figs. 1–3). Furthermore, we have recovered the presence of MTP and Vtg genes in several bilaterians leeches (*Helobdella robusta*), limpets (*Lottia gigantean*), sea urchins (*S. purpuratus*), cephalochordates (*B. floridae*), and tunicates (*Ciona intestinalis*) (Fig. 3). The domain structures of these potentially newly identified bilaterian LLTP genes were compared and are congruent with the domain classification by Avarre and colleagues [1] (Fig. 4), suggesting that bilaterians Apos contain LPD_N, DUF1943, DUF1081 and with or without vWD superfamily domains, whereas the Vtgs contain LPD_N, DUF1943, and vWD superfamily domains, and the MTPs contain the LPD_N domain only.

Contrary to the situation in bilaterians, LLTP genes in non-bilaterian lineages are far less well-characterized. To date, there is only one LLTP gene reported outside bilaterians – where Hayakawa and colleagues [10] propose that a coral sequence (*Galaxea fascicularis*) belongs to the Vtg superfamily. In order to trace the origin of LLTP genes from the less well-characterized non-bilaterians, we constructed phylogenetic trees in an attempt to classify the newly identified LLTP genes recovered in the genomes of non-bilaterians. However, the Bayesian and Maximum-Likelihood analyses did not confidently assign the non-bilaterian putative LLTP genes to any groups included in our analyses (Figs. 1 and 2). This phenomenon is relatively well known as resolving the phylogenetic positions of non-bilaterians genes have occasionally proven to be difficult, which may be attributed to their early divergence from bilaterians and possible elevated rate of evolution (e.g. some homeobox gene families in cnidarians, [15]). Therefore, we have decided to adopt another strategy by using the more robust domain classification for assigning the non-bilaterian LLTP genes.

Through implementation of the conserved domain search for the *G. fascicularis* sequence in the NCBI database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), we identified that this sequence contains LPD_N, DUF1943, and vWD superfamily domains (Fig. 4, Supplementary data 2). This combination of domain structure is expected for the Vtg sequences. To further test whether other LLTP genes could be contained in other cnidarians, we examined the domain structure of LLTP sequences contained within the genome of sea anemone *Nematostella vectensis*. This genome contains LLTP sequences only with similar domain structures to the Vtg (Fig. 4, Supplementary data 3–5). These observations suggest that Vtg is the only LLTP gene family contained among the cnidarian lineages.

Based on the only reported non-bilaterian LLTP gene in coral *G. fascicularis* as a Vtg [10], it is generally postulated that the function of Vtg as the egg-yolk precursor was already present before the divergence of the cnidarians and bilaterians. In the genome of placozoan *Trichoplax adhaerens*, we have unexpectedly identified Vtg orthologues based on domain structure classification (Fig. 4). Similar to the bilaterian Vtg sequences, these sequences do not contain the DUF1943 domain but other domains together with the LPD_N domain (Fig. 4, Supplementary data 6). This result allows us to confidently assign these placozoan sequences as the first *bona fide* Vtg identification outside of the cnidarians, and suggests that the Vtg family was present before the split of cnidarians and

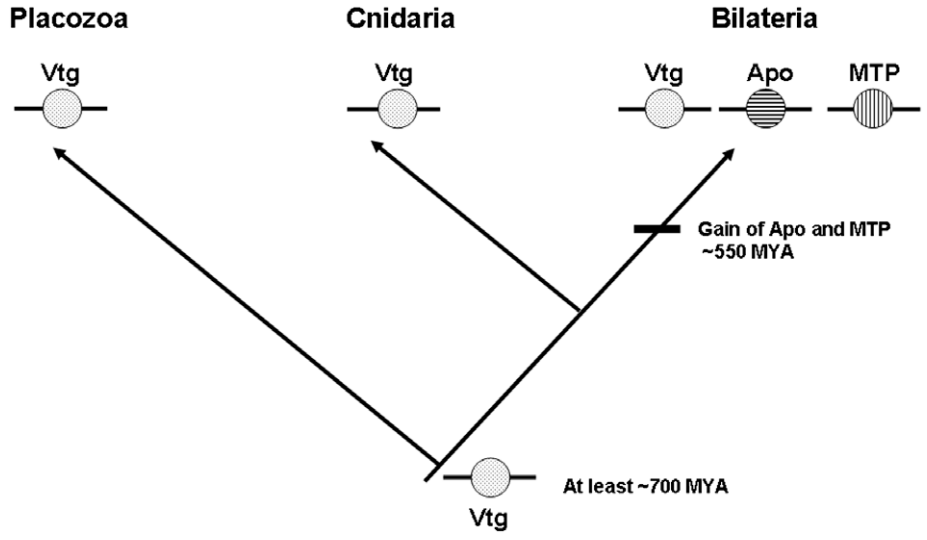

Fig. 5. Proposed model for LLTP genes diversification in animals. The most ancestral LLTP gene is Vtg and existed deep in animal ancestry, which could possibly date back to at least 700 million years ago. Domains loss of Vtg then in the last common ancestor of bilaterians, and gave rise to Apo and MTP. (For details, please refer to text).

bilaterians. This finding suggests that the origin of Vtg is far more ancient than previously thought.

Our classification of the placozoan sequences as Vtg orthologue subsequently raised the question, what is the reason that a well known reproductive gene (Vtg) is now contained within an asexual organism? We speculate that the current function of Vtg in the placozoans could have been related to their feeding modes or behavior, for example, storing lipoproteins as food reserves inside the animals which contribute to their asexual reproduction later in life. Nevertheless, expression and functional analyses in *T. adherens* will be needed to confirm/refute such a hypothesis. For instance, carrying out RNAi knockdown of Vtg expression in *T. adherens* could result in reproduction/fission inability. Regardless of the debatable timing for Vtg function as an egg-york precursor in metazoans (whether the function becomes "derived" in *T. adherens*, or the event originated at the ancestral bilaterians), we speculate that Vtg in placozoans *T. adherens* should now have different functions to the Vtg among extant bilaterians.

To further test whether the origin of LLTP genes could be traced prior to the divergence of placozoans, we also BlastP and TBlastN searched the genome of the choanoflagellate *Monosiga brevicollis* [6]. However, we were not able to identify any LLTP genes in the choanoflagellate genome.

In contrast to the current hypothesis based on the only cnidarian "Vtg" sequence, we propose that Vtg gene is present even deeper in the animal ancestor (Fig. 5). The Vtg gene is putatively the most ancestral LLTP genes that existed in the last common ancestor of the majority of animals at least 700 million years ago. Loss of Vtg domains may then have occurred in the last common ancestor of bilaterians at about 550 million years ago, giving rise to the Apo and MTP genes (Fig. 5). The majority of LLTP genes were most likely retained in the animals as a result of a strong selective force due either to developmental, metabolic or/and reproductive functions. Finally, neofunctionalisation and subfunctionalisation of LLTP genes could also have occurred in different animal lineages as discussed elsewhere [1,2].

Acknowledgements

The authors would like to thank Matthew Ronshaugen for helping with the initial phylogenetic analysis. JHHL would also like to thank David Ferrier and Peter Holland for advice and support.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.02.056.

References

[1] Avarre, J.C., Lubzens, E. and Babin, P.J. (2007) Apolipocrustacein, formerly vitellogenin, is the major egg yolk precursor protein in decapod crustaceans and is homologous to insect apolipophorin II/I and vertebrate Apo B. BMC Evol. Biol. 7, 3.

[2] Babin, P.J. and Gibbons, G.F. (2008) The evolution of plasma cholesterol: direct utility or a "spandrel" of hepatic lipid metabolism? Prog. Lipid Res. 48 (2), 73–91.

[3] Wallace, R.A., Walker, S.L. and Hauschka, P.V. (1967) Crustacean lipovitellin Isolation and characterization of the major high-density lipoprotein from the eggs of decapods. Biochemistry 6, 1582–1590.

[4] Byrne, B.M., Gruber, M. and Ab, G. (1989) The evolution of egg yolk proteins. Prog. Biophys. Mol. Biol. 53, 33–69.

[5] Wilder, M.N., Subramoniam, T. and Aida, K. (2002) Yolk Proteins of Crustacea in: Progress in Vitellogenesis (Adiyodi, K.G., Adiyodi, R.G. and Enfield, N.H., Eds.), Sciences Publishers, USA. pp. 131–174.

[6] King, N., Westbrook, M.J., Young, S.L., Kuo, A., Abedin, M., Chapman, J., Fairclough, S., Hellsten, U., Isogai, Y., Letunic, I., Marr, M., Pincus, D., Putnam, N., Rokas, A., Wright, K.J., Zuzow, R., Dirks, W., Good, M., Goodstein, D., Lemons, D., Li, W., Lyons, J.B., Morris, A., Nichols, S., Richter, D.J., Salamov, A., Sequencing, J.G., Bork, P., Lim, W.A., Manning, G., Miller, W.T., McGinnis, W., Shapiro, H., Tjian, R., Grigoriev, I.V. and Rokhsar, D. (2008) The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451 (7180), 783–788.

[7] Putnam, N.H., Srivastava, M., Hellsten, U., Dirks, B., Chapman, J., Salamov, A., Terry, A., Shapiro, H., Lindquist, E., Kapitonov, V.V., Jurka, J., Genikhovich, G., Grigoriev, I.V., Lucas, S.M., Steele, R.E., Finnerty, J.R., Technau, U., Martindale, M.Q. and Rokhsar, D.S. (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317 (5834), 86–94.

[8] Putnam, N.H., Butts, T., Ferrier, D.E., Furlong, R.F., Hellsten, U., Kawashima, T., Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.K., Benito-Gutiérrez, E.L., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Kuroki, Y., Lindquist, E., Lucas, S., Osoegawa, K., Pennacchio, L.A., Salamov, A.A., Satou, Y., Sauka-Spengler, T., Schmutz, J., Shin-I, T., Toyoda, A., Bronner-Fraser, M., Fujiyama, A., Holland, L.Z., Holland, P.W., Satoh, N. and Rokhsar, D.S. (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453 (7198), 1064–1071.

[9] Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., Kawashima, T., Kuo, A., Mitros, T., Salamov, A., Carpenter, M.L., Signorovitch, A.Y., Moreno, M.A., Kamm, K., Grimwood, J., Schmutz, J., Shapiro, H., Grigoriev, I.V., Buss, L.W., Schierwater, B., Dellaporta, S.L. and Rokhsar, D.S. (2008) The Trichoplax genome and the nature of placozoans. Nature 454 (7207), 955–960.

[10] Hayakawa, H., Andoh, T. and Watanabe, T. (2006) Precursor structure of egg proteins in the coral *Galaxea fascicularis*. Biochem. Biophys. Res. Commun. 344 (1), 173–180.

[11] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) The CLUSTAL_X windows interface. flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

[12] Ronquist, F. and Huelsenbeck, J.P. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

[13] Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52 (5), 696–704.

[14] Abascal, F., Zardoya, R. and Posada, D. (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21 (9), 2104–2105.

[15] Hui, J.H., Holland, P.W. and Ferrier, D.E. (2008) Do cnidarians have a ParaHox cluster? Analysis of synteny around a Nematostella homeobox gene cluster. Evol. Dev. 10 (6), 725–730.